Emergence of skew distributions in controlled growth processes

Segun Goh,[1] H. W. Kwon,[1] M. Y. Choi,[1, *] and J.-Y. Fortin[2]

[1]*Department of Physics and Astronomy and Center for Theoretical Physics,*
Seoul National University, Seoul 151-747, Korea
[2]*Institut Jean Lamour, Département de Physique de la Matière et des Matériaux,*
Nancy Université, BP 70239 F-54506 Vandoeuvre les Nancy Cedex, France

Starting from a master equation, we derive the evolution equation for the size distribution of elements in an evolving system, where each element can grow, divide into two, and produce new elements. We then probe general solutions of the evolution equation, to obtain such skew distributions as power-law, log-normal, and Weibull distributions, depending on the growth or division and production. Specifically, repeated production of elements of uniform size leads to power-law distributions, whereas production of elements with the size distributed according to the current distribution as well as no production of new elements results in log-normal distributions. Finally, division into two, or binary fission, bears Weibull distributions. Numerical simulations are also carried out, confirming the validity of the obtained solutions.

I. INTRODUCTION

The central limit theorem is one of the most important theorems in probability theory. It states that the sum of a large number of independent random variables has a Gaussian distribution, which is symmetric. In reality, however, there exist many phenomena to which the central limit theorem is not applicable. In most of those cases, asymmetric skew distributions arise, characterized by long tails on one side; prototypic examples include power-law, log-normal, and Weibull distributions. In contrast to the Gaussian distribution, of which the origin is well known, there still lacks appropriate understanding of the origin of skew distributions.

The power-law distribution is observed in various physical, biological, economical, or social systems and there are extensive studies attempting to understand its properties [1–3]; *inter alia*, scale invariance is the most important property. While Zipf's law [4] and Pareto's law [5] provide well-known examples of the power-law distribution, the Yule process [6], Gibrat's law [7], or preferential attachment [8] has been proposed as the mechanism for the power-law distribution. Related to the power-law distribution, the log-normal distribution has been used as an alternative of the former [9–11]. Unlike the Gaussian distribution, the log-normal distribution emerges from multiplication (rather than addition) of a large number of independent random variables [12]. Thus geometric Brownian motion or the Black-Scholes equation is related to the log-normal distribution [13]. Interrelations between the power-law distribution and the log-normal one have drawn a great deal of attention in literature and the concept of the double Pareto distribution has also been proposed [14, 15]. The Weibull distribution [16] is also one of the widely applied distribution functions in a variety of systems [11, 17–19]. Emergence of the Weibull distribution has been attributed to diverse origins on the case-by-case basis, including the extreme value statistics [20, 21], fractal cracking [17], Fickian diffusion in the fractal and non-fractal spaces [22, 23], and the branching process [24]. Interrelations between the Weibull distribution and the log-normal one have also been studied and the similarity between them has been elucidated [25]. There are also works which analyzed a variety of time series data, e.g., from respiratory dynamics, electroencephalogram (EEG), heartbeats, and stock price records, and probed the asymmetry of distributions [26–29].

Even though some relations between these skew distributions are available, there still lacks theoretical understanding of how the skew distributions appear in evolving systems. Recently, a general framework has been proposed, based on the master equation describing evolving systems, to elucidate the emergence of skew distributions [30]. Following this approach, we begin with the same master equation, to obtain the time evolution equation for the size distribution in evolving systems, which provides refinements and extensions. It is found that such skew distributions indeed emerge, depending on the growth, division, and production. In particular, the detailed conditions for power-law, log-normal, and Weibull distributions are clarified. In addition, numerical simulations are also carried out extensively; results are shown to agree well with the analytical ones.

There are five sections in this paper: In Sec. II we introduce the master equation approach to the proliferation process, including growth, division, and production. Section III is devoted to solving the evolution equation for various cases. Such skew distributions as power-law, log-normal, and Weibull distributions are obtained in the systems with uniform production, self-production, and binary fission. Section IV presents the results of numerical simulations, which confirm the analytical results in Sec. III. Finally, a summary together with a discussion of the results is given in Sec. V.

* mychoi@snu.ac.kr

II.   MASTER EQUATION

We consider a system of N elements, the ith of which is characterized by the size x_i $(i = 1, \ldots, N)$. The con-figuration of the system is specified by the sizes of all elements, $\{x_1, \ldots, x_N\}$. The probability $P(x_1, \ldots, x_N; t)$ for the system to be in configuration $\{x_1, \ldots, x_N\}$ at time t is governed by the master equation

$$\frac{d}{dt}P(x_1, \ldots, x_N; t) = \sum_{i=1}^{N} \int dx_i' \left[\omega(x_i' \to x_i) P(x_1, \ldots, x_i', \ldots, x_N; t) - \omega(x_i \to x_i') P(x_1, \ldots, x_N; t) \right], \qquad (1)$$

where $\omega(x_i \to x_i')$ is the transition rate for the ith element to change its size from x_i to x_i'. Here we consider the case that the size changes by the amount proportional to the current size. The transition rate then takes the form

$$\omega(x \to x') = \lambda \delta(x' - (1+b)x) \qquad (2)$$

with the (mean) growth rate λ and the growth factor b [30].

We are interested in the size distribution of the system $f(x, t)$, related to the probability $P(x_1, \ldots, x_N; t)$ via:

$$f(x, t) = \frac{1}{N} \int d^N x \sum_{i=1}^{N} \delta(x_i - x) P(x_1, \ldots, x_N; t), \qquad (3)$$

where $\int d^N x \equiv \int_0^\infty dx_1 \cdots \int_0^\infty dx_N$. The evolution equation for the size distribution $f(x, t)$ can be obtained straightforwardly from Eq. (1). In the case that the total number of elements remains constant, the evolution equation for $f(x, t)$ reads

$$\frac{\partial}{\partial t} f(x, t) = -\lambda f(x, t) + \frac{\lambda}{1+b} f\left(\frac{x}{1+b}, t\right). \qquad (4)$$

Note the scale invariance of Eq. (4): Changing the scale according to $x \to cx$ and multiplying Eq. (4) by c for normalization, we obtain Eq. (4) in the form

$$\frac{\partial}{\partial t} cf(cx, t) = -\lambda cf(cx, t) + \lambda \frac{c}{1+b} f\left(\frac{cx}{1+b}, t\right). \qquad (5)$$

This manifests that for $f(x, t)$ satisfying Eq. (4), $cf(cx, t)$ also satisfies Eq. (4), thus providing a solution.

On the other hand, when the total number of elements varies with time, i.e., $N = N(t)$, it is not allowed to interchange the order of d/dt and $\int d^N x$. To circumvent this difficulty, we deal with the time derivative of $N(t)f(x, t)$:

$$\frac{\partial}{\partial t}[N(t)f(x, t)] = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\int d^{N+\Delta N} x \sum_{i=1}^{N+\Delta N} \delta(x_i - x) P(x_1, \ldots, x_{N+\Delta N}; t+\Delta t) - \int d^N x \sum_{i=1}^{N} \delta(x_i - x) P(x_1, \ldots, x_N; t) \right]$$

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\{ \int d^N x \sum_{i=1}^{N} \delta(x_i - x) \left[P(x_1, \ldots, x_N; t+\Delta t) - P(x_1, \ldots, x_N; t) \right] \right.$$

$$\left. + \int dx_{N+1} \cdots dx_{N+\Delta N} \sum_{i=N+1}^{N+\Delta N} \delta(x_i - x) P(x_{N+1}, \ldots, x_{N+\Delta N}; t+\Delta t) \right\}, \qquad (6)$$

where it has been noted that $\int dx_{N+1} \cdots dx_{N+\Delta N} P(x_1, \ldots, x_{N+\Delta N}; t+\Delta t) = P(x_1, \ldots, x_N; t+\Delta t)$ and $\int d^N x \, P(x_1, \ldots, x_{N+\Delta N}; t+\Delta t) = P(x_{N+1}, \ldots, x_{N+\Delta N}; t+\Delta t)$.

The first term on the right-hand side (the second line) of Eq. (6) contains the information only on the existing elements (up to N) and is exactly the same as the case for fixed N, thus resulting in

$$\text{(1st term)} = N(t) \left[-\lambda f(x, t) + \frac{\lambda}{1+b} f\left(\frac{x}{1+b}, t\right) \right]. \quad (7)$$

The second term (the third line) of Eq. (6) is involved with the information on the newly produced elements and takes the form:

$$\text{(2nd term)} = \dot{N}(t) g(x, t) = r N(t) g(x, t), \qquad (8)$$

where $g(x, t)$, define to be

$$g(x,t) \equiv \frac{1}{\Delta N} \int dx_{N+1} \cdots dx_{N+\Delta N} \sum_{i=N+1}^{N+\Delta N} \delta(x_i - x) P(x_{N+1}, \ldots, x_{N+\Delta N}; t), \qquad (9)$$

corresponds to the size distribution of newly produced elements at time t and r is the mean production rate. Namely, it is assumed that each element tends to produce a new one with rate r, which in turn leads the total number of elements to increase in proportion to the current number: $\dot{N} = rN$. Here care needs to be given to the limit $\Delta t \to 0$ since ΔN is discrete, taking only integer values. The expression $\Delta N = rN\Delta t$ is valid as long as $N\Delta t$ is finite, so that ΔN remains discrete. Accordingly, Δt can be reduced down to $(rN)^{-1}$, which corresponds to the time interval required to produce a new element ($\Delta N = 1$). This vanishes as N grows large, thus allowing $\lim_{\Delta t \to 0} \Delta N/\Delta t = \dot{N}$, etc. With the help of Eqs. (6)-(9) and

$$\frac{\partial}{\partial t} f(x,t) = \frac{1}{N}\frac{\partial}{\partial t}[Nf(x,t)] - \frac{\dot{N}}{N} f(x,t), \qquad (10)$$

the time evolution equation for the size distribution finally obtains the form:

$$\frac{\partial f(x,t)}{\partial t} = -(r+\lambda)f(x,t) + \frac{\lambda}{1+b} f\left(\frac{x}{1+b}, t\right) + rg(x,t). \qquad (11)$$

III. ASYMPTOTIC SOLUTIONS

In this section we probe analytically the solutions of the evolution equations (4) and (11). In the case of no production ($r = 0$), the time evolution of $f(x,t)$ is described by Eq. (4). Changing the variable from x to $\ln x \equiv X$, we write the distribution function for X: $F(X,t) = e^X f(e^X, t)$. Correspondingly, Eq. (4) becomes

$$\frac{\partial}{\partial t} F(X,t) = -\lambda F(X,t) + \lambda F(X-a,t), \qquad (12)$$

with $a \equiv \ln(1+b)$. Given the initial distribution $f(x,0) = \delta(x - x_0)$ or equivalently, $F(X,0) = \delta(X - X_0)$, $F(X,t)$ vanishes unless $X = X_n \equiv X_0 + an$ for any integer n (note that this corresponds simply to $x_n = x_0 e^{an} = (1+b)^n x_0$). With the definition $F_n(t) \equiv F(X_0 + an, t)$, Eq. (12) reads

$$\dot{F}_n(t) = -\lambda F_n(t) + \lambda F_{n-1}(t), \qquad (13)$$

which bears the Poisson distribution

$$F_n(t) = \frac{1}{n!} (\lambda t)^n e^{-\lambda t} \qquad (14)$$

for $n \geq 0$ and $F_n(t) = 0$ otherwise [31].

This result can be also recovered by considering the Fourier transform $\tilde{F}(k,t)$ of Eq. (13), which satisfies

$$\partial_t \tilde{F}(k,t) = \lambda(e^{-ika} - 1)\tilde{F}(k,t). \qquad (15)$$

The solution $\tilde{F}(k,t) = \exp[-\lambda(1 - e^{-ika})t]\tilde{F}(k,0)$ of Eq. (15) leads directly to

$$F(X,t) = \int dX' G_F(X, X'; t) F(X', 0)$$
$$= G_F(X, X_0; t), \qquad (16)$$

where $G_F(X, X'; t) = (2\pi)^{-1} \int dk \exp[ik(X-X') - \lambda(1 - e^{-ika})t)]$ and the initial condition $F(X', 0) = \delta(X' - X_0)$ has been noted.

When $t = 0$, we have simply $F(X,0) = G_F(X, X_0; t = 0) = \delta(X - X_0)$. For large values of λt, the function G_F is dominated by small values of k; in this limit, using the saddle-point analysis, we find that Eq. (14) is equivalent to the normal distribution:

$$F(X, t|X_0) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left[-\frac{(X - X_0 - \mu_t)^2}{2\sigma_t^2}\right], \qquad (17)$$

where $\mu_t = a\lambda t$, $\sigma_t = a\sqrt{\lambda t}$, and $F(X, t|X_0)$ stands for the distribution $F(X,t)$ subject to the initial distribution $F(X,0) = \delta(X - X_0)$. The same result is obtained when the growth factor b and hence a are small; this allows us to expand the term $\exp(-ika)$ in the exponential integrand and to perform the gaussian integration. We may also expand G_F directly as the series

$$G_F(X, X'; t) = \sum_{n \geq 0} \int \frac{dk}{2\pi} e^{ik(X-X') - \lambda t} \frac{(\lambda t)^n}{n!} e^{-ikan}$$
$$= \sum_{n \geq 0} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \delta(X - X' - an)$$
$$= \sum_{n \geq 0} F_n(t) \delta(X - X' - an), \qquad (18)$$

which is nothing but the expansion in terms of the Poisson distribution at given time.

In consequence, the original size distribution assumes the log-normal distribution in the long-time limit:

$$f(x, t|x_0) = \frac{1}{\sqrt{2\pi}\sigma_t x} \exp\left[-\frac{(\ln x - \ln x_0 - \mu_t)^2}{2\sigma_t^2}\right], \qquad (19)$$

where $f(x, t|x_0)$ denotes the size distribution $f(x,t)$ subject to the initial distribution $f(x,0) = \delta(x - x_0)$. Given an arbitrary initial distribution $f(x,0)$, the solution thus takes the form

$$f(x,t) = \int dx' G_f(x, x'; t) f(x', 0) \qquad (20)$$

with

$$G_f(x, x'; t) = \frac{1}{x} G_F(\ln x, \ln x'; t). \qquad (21)$$

Equation (20), like Eq. (4), is invariant under the scale transformation $x \to cx$ and $x' \to cx'$.

We now consider the case $r \neq 0$. Expanding the second term on the right-hand side of Eq. (11):

$$f\left(\frac{x}{1+b}, t\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\left(\frac{b}{1+b}\right)^n x^n \frac{\partial^n}{\partial x^n} f(x,t),$$

$$(22)$$

we write Eq. (11) in the compact form

$$\mathcal{L} f(x,t) = r g(x,t), \qquad (23)$$

where the linear operator $\mathcal{L}$ is given by

$$\mathcal{L} \equiv \frac{\partial}{\partial t} + r + \lambda - \frac{\lambda}{1+b}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\left(\frac{b}{1+b}\right)^n x^n \frac{\partial^n}{\partial x^n}.$$

$$(24)$$

Since the solution for $r = 0$, Eq. (19), reduces to the Dirac delta function $\delta(x - x_0)$ in the limit $t \to 0$, Green's function $G(x, x'; t, t') = G(x, x'; t - t')$ for $\mathcal{L}$ can be written in the form

$$G\left(x, x'; t-t'\right) = G_f\left(x, x'; t-t'\right) e^{-r\left(t-t'\right)}\theta\left(t-t'\right), \quad (25)$$

with the Heaviside step function $\theta(t - t')$.

This leads to the general solution of Eq. (11) in the form:

$$f(x,t) = e^{-rt}\int dx' \, G_f(x,x';t) \, f(x',0) + r\int_0^t dt' \, e^{-r(t-t')}\int dx' \, G_f(x,x';t-t')g(x',t'), \qquad (26)$$

where Green's function reduces to the log-normal distribution in the long-time limit ($t \to \infty$):

$$G_f(x, x'; t) \longrightarrow \frac{1}{\sqrt{2\pi}\sigma_t x}\exp\left[-\frac{(\ln x - \ln x' - \mu_t)^2}{2\sigma_t^2}\right].$$

The first term on the right-hand side of Eq. (26), which describes the information on the initial distribution, de-

creases exponentially in time; the second term thus becomes dominant in the long time.

A. Uniform-size production

In case that new elements are produced in uniform size x_0, i.e., $g(x,t) = \delta(x - x_0)$, the asymptotic solution in the long-time limit reads

$$f_s(x) \equiv f(x, t \to \infty) = \lim_{t \to \infty} r\int_0^t dt' \, e^{-r(t-t')}G_f(x, x_0; t-t') = \frac{r}{x}\int \frac{dk}{2\pi}\frac{e^{ik\ln(x/x_0)}}{r + \lambda(1 - e^{-ika})}. \qquad (27)$$

The integral over k can be performed with the help of the decomposition $ka = 2\pi n - \theta$, where n is an integer and $0 \le \theta < 2\pi$. Equation (27) then obtains the form of an infinite sum over integers n:

$$f_s(x) = \frac{r}{ax}\sum_n e^{2i(\pi/a)n\ln(x/x_0)}\int_0^{2\pi}\frac{d\theta}{2\pi}\frac{e^{-(i/a)\theta\ln(x/x_0)}}{r + \lambda - \lambda e^{i\theta}}.$$

We notice that the integral over θ is independent of n, and use the Dirac comb identity to write

$$f_s(x) = \frac{r}{ax}\sum_n \delta\left(n - a^{-1}\ln(x/x_0)\right)\int_0^{2\pi}\frac{d\theta}{2\pi}\frac{e^{-(i/a)\theta\ln(x/x_0)}}{r + \lambda - \lambda e^{i\theta}}$$

$$= \frac{r}{ax}\sum_n \delta\left(n - a^{-1}\ln(x/x_0)\right)\int_0^{2\pi}\frac{d\theta}{2\pi}\frac{e^{-in\theta}}{r + \lambda - \lambda e^{i\theta}},$$

where it has been noted that the argument $a^{-1}\ln(x/x_0)$ is constrained by each integer n. The integral over θ can

be transformed into one over a unit circle in a complex plane:

$$\int_0^{2\pi}\frac{d\theta}{2\pi}\frac{e^{-in\theta}}{r + \lambda - \lambda e^{i\theta}} = \oint \frac{dz}{2i\pi z}\frac{z^{-n}}{r + \lambda - \lambda z},$$

where we have set $e^{i\theta} \equiv z$. This integral vanishes when n is negative, i.e., for $x < x_0$; otherwise Cauchy's residue theorem yields

$$\oint \frac{dz}{2i\pi z}\frac{z^{-n}}{r + \lambda - \lambda z} = \frac{\lambda^n}{(r+\lambda)^{n+1}}$$

for $n \ge 0$. Setting finally $x_n = x_0 e^{an} = x_0(1+b)^n$, we obtain the stationary distribution

$$f_s(x) = \frac{r}{r+\lambda}\sum_{n \ge 0}\left(\frac{x}{x_0}\right)^{-\frac{\ln(1+r/\lambda)}{\ln(1+b)}}\delta(x - x_n). \qquad (28)$$

On the other hand, the stationary solution can be obtained simply by putting $\partial f/\partial t = 0$ and $g(x,t) =$

$\delta(x - x_0)$ in Eq. (11). In this way, we still obtain the power-law distribution for $x > x_0$:

$$f_s(x) \sim x^{-\alpha} \qquad (29)$$

with the exponent $\alpha = 1 + \frac{\ln(1+r/\lambda)}{\ln(1+b)}$, which should be exact in the stationary configuration. Equations (28) and (29) give the same kind of power-law distribution in the long-time limit. It is indeed possible to recover from Eq. (28) the simple stationary power-law solution for small b or a. In such a continuous-growth limit, we also have $r \to 0$, with the ratio r/b and parameter λ kept finite, since $[\ln(1+b)]^{-1}\ln(1+r/\lambda)$ and thus α would diverge otherwise. In this case Eq. (28) may be approximated by replacing the discrete sum over n by an integral over the variable $u = an$ which becomes continuous in the limit $a \to 0$:

$$f_s(x) = \frac{r}{r+\lambda}\frac{1}{x}\sum_{n\geq 0}\left(\frac{x}{x_0}\right)^{-\frac{\ln(1+r/\lambda)}{\ln(1+b)}}\delta(an - \ln(x/x_0))$$

$$\approx \frac{r}{r+\lambda}\frac{1}{ax}\int_0^\infty du\left(\frac{x}{x_0}\right)^{1-\alpha}\delta(u - \ln(x/x_0))$$

$$= \frac{r}{\lambda b x_0}\left(\frac{x}{x_0}\right)^{-\alpha}\theta(x - x_0), \qquad (30)$$

where θ is the Heaviside step function. This normalized power-law distribution has therefore the same exponent $\alpha = 1 + [\ln(1+b)]^{-1}\ln(1+r/\lambda) \approx 1 + r/b\lambda$ as the stationary one given by Eq. (29) in the continuous-growth limit. Accordingly, our general solution given by Eq. (26) demonstrates that the distribution, starting from the Dirac delta distribution, evolves to the power-law distribution in the uniform-size production process.

B. Self-size production

When an element produces a new element of the same size as itself, the size distribution of newly produced elements is identical to that of the system, i.e., $g(x,t) = f(x,t)$. In this case, the size distribution of newly produced elements is not given externally, and Eq. (26) no longer provides the solution of Eq. (11) but makes an integral equation for $f(x,t)$. Remarkably, however, putting $g(x,t) = f(x,t)$ in Eq. (11) reduces the equation precisely to Eq. (4). Namely, the case of self-size production is effectively the same as the case of no production ($r = 0$), except for the increase of the total number of elements due to the production of new elements. In consequence, the size distribution of the system with self-size production is given by Eq. (20). Starting from the initial configuration that all elements have the identical size x_0, the size distribution develops to the log-normal distribution:

$$f(x,t) = \frac{1}{\sqrt{2\pi}\sigma_t x}\exp\left[-\frac{(\ln x - \mu_t)^2}{2\sigma_t^2}\right] \qquad (31)$$

with $\mu_t = \ln(1+b)\lambda t + \ln x_0$ and $\sigma_t = \ln(1+b)\sqrt{\lambda t}$.

C. Binary fission

The binary fission by means of division in size may be interpreted as follows: The size of an element is reduced by the multiplicative factor $\beta < 1$ at rate r and a new element with the size reduced by the multiplicative factor $1 - \beta$ is produced at the same rate r. As a result, the corresponding evolution equation obtains the form

$$\frac{\partial f(x,t)}{\partial t} = -(2r + \lambda)f(x,t) + \frac{\lambda}{1+b}f\left(\frac{x}{1+b},t\right) + \frac{r}{\beta}f\left(\frac{x}{\beta},t\right) + \frac{r}{1-\beta}f\left(\frac{x}{1-\beta},t\right), \qquad (32)$$

which describes both growth and reduction of the size due to binary fission. Although it can be easily extended further to describe more complicated processes, we consider only one each kind of size-growth process and size-reduction process, which suffices for our study here. Specifically, we suppose that the size of an element remains constant while the element gives birth to an offspring of the size proportional to its size with the multiplicative factor $1 - \beta$:

$$\frac{\partial f(x,t)}{\partial t} = -(r + \lambda)f(x,t) + \frac{\lambda}{1+b}f\left(\frac{x}{1+b},t\right) + \frac{r}{1-\beta}f\left(\frac{x}{1-\beta},t\right), \qquad (33)$$

with $b > 0$ and $0 < \beta < 1$. Note that, in the half-size division process ($\beta = 1/2$), Eqs. (32) and (33) become identical.

In contrast to the cases of uniform-size production and self-size production, the size distribution of newly produced elements depends on that of the system and does not cancel out, respectively. Making use of the scaling symmetry of Eq. (33), one is tempted to take a log-normal or a power-law distribution as an ansatz solution. However, neither the log-normal distribution nor the power-law distribution can make a stable solution, as manifested easily by substitution. This may be inferred from the observation that the linear operator $\mathcal{L}$ for Eq. (33) cannot be cast into the same form as that for Eq. (24). Therefore we introduce another distribution observing the scaling symmetry, namely, the Weibull distribution:

$$f(x,t) = \frac{\gamma}{x}\left(\frac{x}{\eta}\right)^{\gamma}\exp\left[-\left(\frac{x}{\eta}\right)^{\gamma}\right] \qquad (34)$$

with the shape parameter γ and the scale parameter η. There exist other distribution functions with the same symmetry, for example, the generalized exponential distribution and the gamma distribution; they also fail to provide asymptotic solutions of Eq. (33).

Henceforth we denote $y \equiv \ln(x/\eta)$, $u \equiv 1-(1+b)^{-\gamma}$, and $v \equiv 1-(1-\beta)^{-\gamma}$ for convenience. Expanding $(x/\eta)^{\gamma}$ at $\gamma = 0$,

$$\left(\frac{x}{\eta}\right)^{\gamma} = \sum_{k=0}^{\infty}\frac{1}{k!}(\gamma y)^k, \qquad (35)$$

we write the time derivative of the Weibull distribution in the form

$$\frac{1}{f}\frac{\partial f}{\partial t} = \frac{\dot{\gamma}}{\gamma} + \left(\dot{\eta}\frac{\gamma}{\eta} - \dot{\gamma}y\right)\sum_{k=1}^{\infty}\frac{1}{k!}(\gamma y)^k, \qquad (36)$$

while the right-hand side of Eq. (33) divided by $f(x,t)$ becomes

$$\frac{1}{f}(\text{rhs}) = -(\lambda + r) + \lambda(1-v)e^v + r(1-u)e^u + \lambda(1-v)e^v\sum_{k=1}^{\infty}\frac{1}{k!}B_k(v)(\gamma y)^k + r(1-u)e^u\sum_{k=1}^{\infty}\frac{1}{k!}B_k(u)(\gamma y)^k \qquad (37)$$

with the Bell polynomial given by [32]

$$B_k(z) \equiv e^{-z}\sum_{n=0}^{\infty}\frac{n^k}{n!}z^n. \qquad (38)$$

Comparison of Eqs. (36) and (37) gives, for small γ,

$$\dot{\gamma} = -\frac{\gamma^3}{2}\left\{\lambda\left[\ln(1+b)\right]^2 + r\left[\ln(1-\beta)\right]^2\right\} + O(\gamma^4) \qquad (39)$$

to the zeroth order in y, and

$$\dot{\eta} = \eta\left[\lambda\ln(1+b) + r\ln(1-\beta)\right] + O(\gamma), \qquad (40)$$

to the nth order in y for all $n \geq 1$, where the expansion of v about $\gamma = 0$,

$$v \equiv 1 - \left(\frac{1}{1+b}\right)^{\gamma} = -\sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\left[\gamma\ln(1+b)\right]^k, \qquad (41)$$

and similar expansion of u (with $-\beta$ in the place of b) have been used. To the leading order, the solutions of Eqs. (39) and (40) under the initial conditions $\gamma(0) = \gamma_0$ and $\eta(0) = \eta_0$ read

$$\gamma(t) = \frac{\gamma_0}{\sqrt{1 + C\gamma_0^2 t}} \qquad (42a)$$

$$\eta(t) = \eta_0\, e^{Dt}, \qquad (42b)$$

where $C \equiv \lambda\left[\ln(1+b)\right]^2 + r\left[\ln(1-\beta)\right]^2$ and $D \equiv \lambda\ln(1+b) + r\ln(1-\beta)$. For this to be valid, one should

have sufficiently small γ and y (or large η). In the asymptotic limit the conditions of small γ and large η require both C and D to be positive. Note that while C is always positive, D is positive only for $r/\lambda < -\ln(1+b)/\ln(1-\beta)$. In conclusion, for $r/\lambda < -\ln(1+b)/\ln(1-\beta)$, the Weibull distribution in Eq. (34) is asymptotically exact, where the shape parameter γ decreases algebraically in time and the scale parameter η grows exponentially in time.

D. Autocorrelations

In the case of an asymmetric distribution of variables, there can exist power-law correlations in the variables for suitably adjusted parameters [33]. To probe the possibility of long-time correlations in the evolution governed by Eq. (11), we consider the autocorrelation function at stationarity:

$$C(t) \equiv \frac{\langle x(\tau)x(\tau+t)\rangle - \langle x(\tau)\rangle\langle x(\tau+t)\rangle}{\sqrt{\langle x^2(\tau)\rangle - \langle x(\tau)\rangle^2}\sqrt{\langle x^2(\tau+t)\rangle - \langle x(\tau+t)\rangle^2}} \qquad (43)$$

where $\langle x(t)\rangle \equiv \int dx\, f(x,t)x$ and $\langle x(\tau)x(\tau+t)\rangle \equiv \int dx\, dx'\, G(x,x';t)f(x',\tau)xx'$ with appropriate Green's function $G(x,x';t)$ and τ being large enough for the system to be in the stationary state.

It is straightforward to compute Eq. (43), which turns out to exhibit exponentially decaying behavior in the long-time limit. Specifically, in the case of the self-size production or no production ($r = 0$), Green's function

reduces simply to Eq. (21): $G(x, x'; t) = G_f(x, x'; t)$, and the resulting autocorrelation function reads

$$C(t) \sim \exp\left(-\frac{b^2\lambda}{2}t\right). \tag{44}$$

For the branching process, Green's function is given by

$$G(x, x'; t) = e^{-rt}G_f(x, x'; t) + r\int_0^t dt' e^{-r(t-t')}G_f(x, (1-\beta)x'; t-t'), \tag{45}$$

which leads to

$$C(t) \sim \exp\left(-\frac{b^2\lambda + \beta^2 r}{2}t\right). \tag{46}$$

Finally, when there is a source producing new elements of uniform size, Green's function takes the form

$$G(x, x'; t) = e^{-rt}G_f(x, x'; t) + r\int_0^t dt' e^{-r(t-t')}G_f(x, x_0; t-t') \tag{47}$$

where x_0 is the size of newly produced elements. In this case the autocorrelation function depends on the production rate r: For $r < 2b\lambda + b^2\lambda$, it reads

$$C(t) \sim \exp\left(-\frac{r + b^2\lambda}{2}t\right) \tag{48}$$

whereas for $r \geq 2b\lambda + b^2\lambda$, we have

$$C(t) \sim \exp\left[(-r + b\lambda)t\right]. \tag{49}$$

In all cases, the autocorrelations decay exponentially in time. We have also carried out numerical simulations to compute autocorrelation functions, to obtain results in full agreement with the analytical results listed above. It is thus concluded that long-time autocorrelations do not exist in the growth process considered here. This apparently implies that skew distributions do not necessarily result from long-time autocorrelations of the data variables.

IV. NUMERICAL SIMULATIONS

We carry out numerical simulations to check the analytical results in the preceding section. The master equation (1) is integrated directly to give the probability distribution for no-production, uniform-size production, self-size production, and half-size division processes. At each time step given by an integer multiple of Δt in simulations, a trial move is attempted according to the growth probability $\Lambda \equiv \lambda\Delta t$ and the production probability $R \equiv r\Delta t$. To access the continuous-time case, Δt should be taken to be sufficiently small. Here we choose
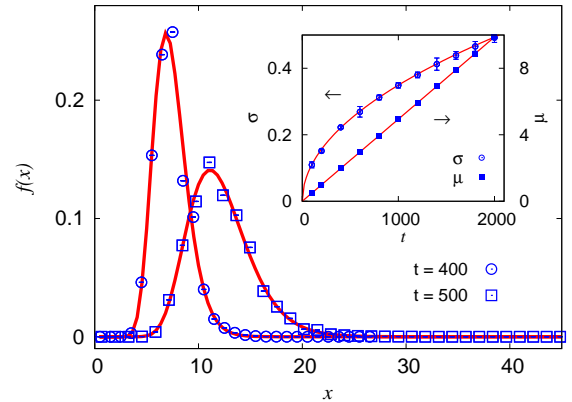

FIG. 1. Evolution of the size distribution in the case of no production, $r = 0$. Circles and squares plot simulation data for $\lambda = 0.2$ and $b = 0.025$ at time $t = 400$ and 500, respectively. Solid lines represent the corresponding analytic solutions, which are log-normal distributions with $(\mu, \sigma) = (1.975, 0.2209)$ and $(2.469, 0.2469)$, respectively. The inset shows σ (circles) and μ (squares) versus time t, disclosing excellent agreement between simulation results (symbols) and analytical ones (solid lines). Error bars have been estimated by the standard deviations of the simulation data, and are smaller than the symbols.

$\Delta t = 0.001$ and perform simulations of 10^2 different samples, over which obtained results are averaged. Initially, all elements in each sample are taken to be of the unit size $(x_0 = 1)$ and the number of elements is usually adjusted so that each sample in the final data consists of $N = 10^4$ to 10^6 elements. These simulation conditions have been varied, to give no appreciable differences.
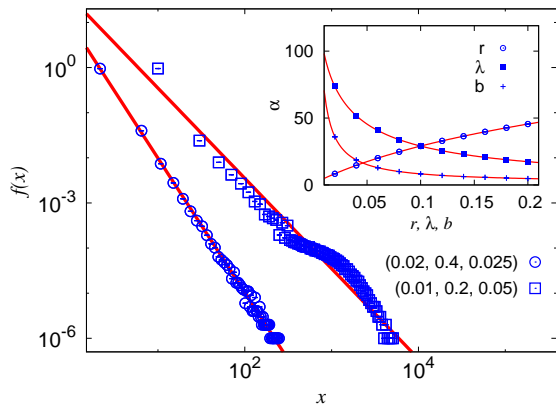
FIG. 2. Evolution of the size distribution in the case of uniform-size production. Circles and squares plot simulation data for $(r, \lambda, b) = (0.02, 0.4, 0.025)$ at time $t = 500$ and $(0.01, 0.2, 0.05)$ at $t = 700$, respectively, in the log-log scale. Solid lines represent stationary solutions, with the slope $-\alpha = -2.976$ and -2, respectively, given by Eq. (29). The inset shows the exponent α of the distribution versus various parameters: Circles display α vs. the production rate r for $\lambda = 0.1$ and $b = 0.025$; squares present α vs. the growth rate λ for $r = 0.1$ and $b = 0.025$, and crosses α vs. the growth factor b for $r = \lambda = 0.1$. Solid lines represent analytical solutions. Error bars estimated by the standard deviations of the simulation data are smaller than the symbols.

When no new elements are produced, the simulation results are shown in Fig. 1, and compared with the analytic solutions. It is observed that the distributions obtained from simulations fit well to the log-normal distributions described by the analytical expression in Eq. (20). The asymptotic solution also gives excellent estimate for the time evolution of the mean and standard deviation parameters (see the inset). This supports strongly the validity of the analysis in Sec. III, including the log-normal kernel in Eq. (19) and the asymptotic general solution in Eq. (26).

Figure 2 shows the simulation results for the case of uniform-size production, where the initial number of elements has been adjusted in such a way that the system ends with $N = 10^6$ elements for both cases, $(r, \lambda, b) = (0.02, 0.4, 0.025)$ and $(0.01, 0.2, 0.05)$. For comparison, analytical results are shown together, manifesting power-law behavior. In particular, the inset exhibits the exponent α of the power-law distribution versus r, λ, and b; there simulation results are compared with the analytical solution given by Eq. (28) or Eq. (29). As expected, excellent agreement is observed.

For the self-size production process $g(x, t) = f(x, t)$, it is expected that simulations give the same results as the no-production cases. To confirm this, we carry out simulations of the system with the production rate $r = 0.01$ and other parameters the same as those in Fig. 1 (i.e., $\lambda = 0.2$ and $b = 0.025$). The resulting distributions at $t = 400$ and 500 (not shown) indeed turn out identical to those in Fig. 1. This also implies that the contribu-
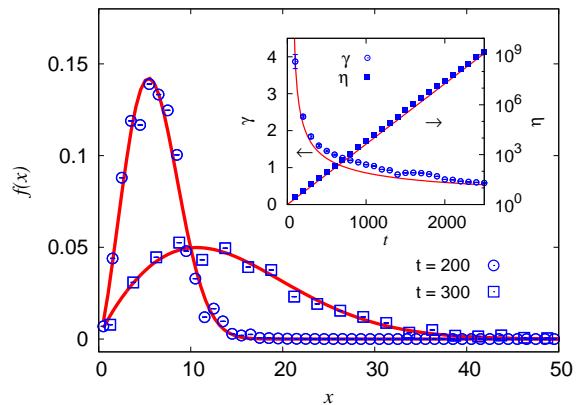

FIG. 3. Evolution of the size distribution in the half-size division process with $r = 0.002$, $\lambda = 0.2$, and $b = 0.05$. Circles and squares plot simulation data at time $t = 200$ and 300, respectively. Solid lines represent the fitted Weibull distribution with $(\gamma, \eta) = (2.380, 6.834)$ for $t = 200$ and $(1.842, 16.35)$ for $t = 300$, respectively. Error bars estimated by the standard deviations are not larger than the symbols. The inset shows γ (circles) and η (squares) versus time t, disclosing good agreement between simulation results (symbols) and analytical ones (solid lines).

tions of the size distribution of newly produced elements to the change in the size distribution of the system are described correctly by Eq. (11).

Lastly, branching processes are simulated and the results for the half-size division process ($\beta = 1/2$) are illustrated in Fig. 3. The resultant distributions fit well to the Weibull distribution. The shape parameter γ and the scale parameter η evolve in time, the agreement of which with the analytical results gets better as time goes on. This suggests that the branching process provides inherent mechanism of the Weibull distribution, the general origin of which has remained unresolved.

V. DISCUSSION

We have established a model for describing the size distribution in evolving systems and probed its solutions in several cases both analytically and numerically. Obtained are skew distributions including power-law, log-normal, and Weibull distributions, depending on the production rate, growth rate, and growth factor. Specifically, power-law distributions are obtained when new elements of uniform size are produced; log-normal distributions emerge in case that each element tends to produce either a new one of the same size as itself or no element at all. Further, branching processes give rise to Weibull distributions, suggestive of the origin of the Weibull distribution.

Emergence of the Weibull distribution in our model needs careful consideration: As the asymptotic expansion does not guarantee uniqueness of the solution, there

may exist other solutions which are distinguishable from our solution by the asymptotic expansion to the leading order. Further, there still lacks understanding of the time evolution of the Weibull distribution, e.g., ranges of the production and growth rates in which the Weibull distribution provides quantitatively better fitting have not searched enough. The predicted time evolution of the shape and scale parameters is not exact, in contrast with the log-normal distribution for the self-size production process. Nevertheless, the physical meaning as to the emergence of the Weibull distribution is suggested convincingly and the accuracy of the asymptotic expansion is also supported to some degree by numerical simulations.

With the insight gained from this study, we may understand skew distributions observed in various systems. For instance, the power-law distribution of the populations in the U.S. cities [34] can be understood as an evolving system, where small towns are created and then grow to larger cities. The number of passengers in an intra-urban subway system [35] may also be treated by this model: The numbers of passengers making trips between two subway stations or using a subway station are indeed observed to follow log-normal or Weibull distributions. The size distribution of beta cells in pancreatic islets,

which has been experimentally measured and theoretically predicted [11], is also described by a log-normal or Weibull distribution. It is expected that there are many additional evolving systems with skew distributions, explained properly by this model.

It is remarkable that ubiquitous emergence of such representative skew distributions as power-law, log-normal, and Weibull distributions is successfully described within a single framework. Applications to real systems exhibiting these skew distributions and more detailed understanding of the emergence of the Weibull distribution are left for further study.

[1] F. Lucchin and S. Matarrese, Phys. Rev. D, **32**, 1316 (1985).
[2] M. Levy and S. Solomon, Physica A, **242**, 90 (1997).
[3] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley, Nature, **423**, 267 (2003).
[4] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, 1949).
[5] V. Pareto, *Cours d'Economie Plitique* (Droz, Geneva, 1896).
[6] G. U. Yule, Phil. Trans. Roy. Soc. London. Ser. B, **213**, 21 (1925).
[7] R. Gibrat, *Les Inégalités Economiques* (Sirey, Paris, 1931).
[8] A.-L. Barabási and R. Albert, Science, **286**, 509 (1999).
[9] W. W. Fleming, D. P. Westfall, I. S. de la Lande, and L. B. Jellett, J. Pharmacol. Exp. Ther., **181**, 339 (1972).
[10] K. Oexle, J. Theor. Biol., **190**, 369 (1998).
[11] J. Jo, M. Y. Choi, and D.-S. Koh, Biophys. J., **93**, 2655 (2007).
[12] E. Limpert, W. A. Stahel, and M. Abbt, BioScience, **51**, 341 (2001).
[13] F. Black and M. Scholes, J. Political Econ., **81**, 637 (1973).
[14] W. J. Reed, Physica A, **319**, 469 (2003).
[15] M. Mitzenmacher, Internet Math., **1**, 226 (2004).
[16] W. Weibull, ASME J. Appl. Mech., **18**, 293 (1951).
[17] T. W. Peterson and M. V. Scotto, Powder Tech., **45**, 87 (1985).
[18] G. R. McDowell, M. D. Bolton, and D. Robertson, J. Mech. Phys. Solids, **44**, 2079 (1996).
[19] V. Papadopoulou, K. Kosmidis, M. Vlachou, and P. Macheras, Int. J. Pharmaceutics, **309**, 44 (2006).
[20] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels, *Statistics of Extremes: Theory and Applications* (Wiley, Chichester, 2004).
[21] E. Bertin and M. Clusel, J. Phys. A, **39**, 7607 (2006).
[22] K. Kosmidis, P. Argyrakis, and P. Macheras, Pharmaceutical Res., **20**, 988 (2003).
[23] K. Kosmidis, P. Argyrakis, and P. Macheras, J. Chem. Phys., **119**, 6373 (2003).
[24] J. Jo, J.-Y. Fortin, and M. Y. Choi, Unpublished.
[25] W. K. Brown and K. H. Wohletz, J. Appl. Phys., **78**, 2758 (1995).
[26] K. Ohashi, L. A. N. Amaral, B. H. Natelson, and Y. Yamamoto, Phys. Rev. E, **68**, 065204 (2003).
[27] C. Braun, P. Kowallik, A. Freking, D. Hadeler, K.-D. Kniffki, and M. Meesmann, Am. J. Physiol. Heart Circ. Physiol., **275**, H1577 (1998).
[28] C. L. Ehlers, J. Havstad, D. Prichard, and J. Theiler, J. Neurosci., **18**, 7474 (1998).
[29] M. Sammon and F. Curley, J. Appl. Physiol., **83**, 975 (1997).
[30] M. Y. Choi, H. Choi, J.-Y. Fortin, and J. Choi, Europhys. Lett., **85**, 30006 (2009).
[31] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Application* (Springer-Verlag, New York, 1996).
[32] S. Roman, *The Umbral Calculus* (Academic Press, Florida, 1984).
[33] B. Podobnik, P. C. Ivanov, V. Jazbinsek, Z. Trontelj, H. E. Stanley, and I. Grosse, Phys. Rev. E, **71**, 025104 (2005).
[34] M. E. J. Newman, Contemp. Phys., **46**, 323 (2005).
[35] K. Lee, W.-S. Jung, J. S. Park, and M. Y. Choi, Physica A, **387**, 6231 (2008).